DERM RAG

by

Anusha Gadgil
Bhavya Kalra
Neeraj Gupta
Raj Acharya

A Project on DERM RAG

Luddy School of Informatics
Indiana University Indianapolis
May 2024

0.1 Introduction

In the field of dermatology, accurately diagnosing skin conditions and their underlying systemic associations is crucial for effective treatment and patient well-being. However, the vast amount of information available in medical literature can be overwhelming for healthcare professionals to navigate efficiently. To address this challenge, we have developed "DermLLaMA," a powerful language model that combines the expertise from the authoritative book "Dermatological Signs of Systemic Disease, 5th Edition" by Callen et al. with state-of-the-art natural language processing techniques.

DermLLaMA is designed to assist healthcare professionals in answering questions related to dermatological signs of systemic diseases. By leveraging the knowledge distilled from the book and the capabilities of large language models, DermLLaMA can provide accurate and contextual responses to user queries. This AI-powered system aims to enhance the decision-making process and improve patient outcomes by making the wealth of information in the book more accessible and user-friendly.

In recent years, the rise of conversational AI models like ChatGPT has revolutionized the way people interact with technology and seek information.These models have demonstrated remarkable language understanding and generation abilities, making them valuable tools for answering questions and providing insights. However, when it comes to medical information, it is crucial to rely on authoritative sources and expert-curated knowledge to ensure the accuracy and reliability of the provided information.

Moreover, it is important to note that while DermLLaMA can be a valuable resource, it should not replace professional medical advice. Small skin symptoms can sometimes be indicative of more serious underlying conditions, such as cancer. In such cases, it is essential for patients to consult with qualified dermatologists or other healthcare professionals for proper diagnosis and treatment.

By combining the expertise from the book "Dermatological Signs of Systemic Disease, 5th Edition" with the power of language models, DermLLaMA aims to bridge the gap between the wealth of medical knowledge and the practical needs of healthcare professionals. This innovative approach to dermatological education and decision support has the potential to transform the way clinicians approach skin-related conditions and their systemic associations.

Chapter 1

Problem Statement and Data Description

The goal is to develop an AI-based system called "DermRAG" that can effectively retrieve queries related to dermatological signs of systemic diseases. This is an important problem as accurately diagnosing skin conditions and their underlying systemic associations is crucial for effective treatment and patient well-being. However, navigating the vast amount of information available in medical literature can be challenging for healthcare professionals. DermRAG aims to bridge this gap by making the knowledge more accessible and user-friendly through the use of large language models.
The dataset consists of two key resources:

1. The book "Dermatological Signs of Systemic Disease, 5th Edition" by Callen et al.1, which provides in-depth information on the cutaneous manifestations of systemic disorders. This authoritative text has become an essential reference for dermatologists, internists, and other healthcare professionals.

2. The "Dermatology Question-Answer Dataset: Skin Disease", a synthetic dataset created by extracting relevant information from the book and formatting it into a structured question-answer format1. This dataset combines the expertise from the book with the capabilities of large language models, specifically the LLaMA-2 model, to develop a powerful AI-based system capable of answering questions related to skin conditions and their systemic associations.

The dataset was constructed by carefully extracting relevant information from the book and formatting it into a structured question-answer format. The LLaMA-2 model was then fine-tuned on this dataset, enabling it to understand and respond to queries about dermatological signs of systemic diseases with high accuracy and contextual understanding.

Chapter 2

Methodology

To create a powerful AI-based system capable of answering questions related to dermatological signs of systemic diseases, we will leverage the knowledge from two key resources: the book "Dermatological Signs of Systemic Disease, 5th Edition" by Callen et al. and the LLaMA-2 large language model. First, we will carefully extract relevant information from the book, including disease descriptions, symptoms, diagnostic criteria, and treatment recommendations. This data will be formatted into a structured question-answer format, creating the "Dermatology Question-Answer Dataset: Skin Disease"1.Next, we will fine-tune the Embedding Model on this dataset using the LlaMA-Index library. The fine-tuning process will involve several steps:

1. Splitting the dataset into training and validation sets using scikit-learn's `train_test_split` function.

2. Generating synthetic questions for each text chunk in the corpus using GPT-3.5-turbo, an LLM provided by OpenAI. Each pair of (generated question, text chunk) will be added to the fine-tuning dataset1.

3. Fine-tuning the Embedding Model on the training set using the —SentenceTransformersFinetuneEngine— from LlaMA-Index1.

4. Evaluating the fine-tuned model's performance on the validation set using two approaches:

   a. A custom hit rate metric that calculates the percentage of queries where the relevant document was retrieved among the top-k results1.

   b. The `InformationRetrievalEvaluator` from the sentence_transformers library, which provides a more comprehensive suite of metrics1.

The fine-tuned model, named "DERMRAG," will be saved for future use and deployment. By combining the expertise from the book with the capabilities of large language models, DERMRAG aims to provide accurate and contextual answers to user queries related to dermatological signs of systemic diseases.This methodology leverages the strengths of both the book and the LLaMA-2 model, creating a synergistic system that can assist healthcare professionals in making more informed decisions and improve patient outcomes.

Chapter 3

Results

The results demonstrate the effectiveness of fine-tuning an embedding model on a synthetic (LLM-generated) dataset for an information retrieval task related to dermatological signs of systemic diseases. The fine-tuning process leveraged the expertise from the book "Dermatological Signs of Systemic Disease, 5th Edition" by Callen et al.1 and the capabilities of large language models, specifically the Embedding model.The fine-tuning process involved several key steps:

1. Extracting relevant information from the book, including disease descriptions, symptoms, diagnostic criteria, and treatment recommendations, and formatting it into a structured question-answer format to create the "Dermatology Question-Answer Dataset: Skin Disease"1.

2. Splitting the dataset into training and validation sets using scikit-learn's `train_test_split` function1.

3. Generating synthetic questions for each text chunk in the corpus using GPT-3.5-turbo, an LLM provided by OpenAI. Each pair of (generated question, text chunk) was added to the fine-tuning dataset1.

4. Fine-tuning the Embedding Model on the training set using the `SentenceTransformersFinetur` from LlaMA-Index1.

   After fine-tuning, the performance of three embedding models was evaluated:

   (a) The proprietary OpenAI embedding
   (b) The open-source BAAI/bge-small-en embedding
   (c) The fine-tuned embedding model ("test_model")

   Two evaluation approaches were used:

   (a) A simple hit rate metric that calculates the percentage of queries where the relevant document was retrieved among the top-k results1.
   (b) The `InformationRetrievalEvaluator` from the sentence_transformers library, which providing comprehensive metrics1.

   The results indicate that the fine-tuned model outperforms others in various metrics:

Table 3.1: Model outcomes

Sr. No.	Model	Hit Rate	MAP
1	Open AI	98.31%	-
2	BAAI	94.59%	87.48%
3	Fine Tuned	95.60%	88.65%

Table 3.2: Model Accuracy

Model	Accuracy@1	Accuracy@3	Accuracy@5	Accuracy@10
Naive BGE/BAAI	81.08%	90.54%	94.59%	96.28%
FIne-tuned BGE/BAAI	81.41%	91.89%	95.60%	97.29%

Table 3.3: Model Precision

Model	Precision@1	Precision@3	Precision@5	Precision@10
Naive BGE/BAAI	81.08%	30.18%	18.91%	9.62%
FIne-tuned BGE/BAAI	81.41%	30.63%	19.12%	9.73%

- The fine-tuned model outperforms other models across various evaluation metrics.
- Hit rate: The fine-tuned model achieves a hit rate of 95.6.
- Score: The fine-tuned model achieves a score of 87.
- Accuracy: The fine-tuned model attains higher Accuracy@1, Accuracy@3, Accuracy@5, and Accuracy@10 scores using both cosine similarity and dot product scoring, indicating superior retrieval of relevant documents in top positions.
- Precision and Recall: The fine-tuned model demonstrates higher Precision@1, Recall@1, Precision@3, and Recall@3 scores, indicating retrieval of a larger proportion of relevant documents in top positions.
- Precision (MAP@10): The fine-tuned model achieves a higher MAP@10, indicating that relevant documents are retrieved at higher positions on average.

References

Literature review:

- Deep Learning in Dermatology: A Systematic Review of Current Approaches, Outcomes, and Limitations
- Hugging Face Documentation
- LLaMA: Open and Efficient Foundation Language Models
- LLaMA2 github Dataset finalized:
  - Dermatology Book: Jeffrey P. Callen, Joseph L. Jorizzo, John J. Zone, Warren Piette - Dermatological Signs of Systemic Disease (2016, Elsevier)
  - Synthetic Data: Dermatology Question-Answer Dataset: Skin Disease

  Code development:
  - Installed required packages for fine-tuning using LlaMA-Index library
  - Defined a function to load corpus from files
  - Split dataset into training and validation sets
  - Generated synthetic questions using GPT-3.5-turbo for fine-tuning
  - Fine-tuned Embedding Model on training set
  - Evaluated performance of OpenAI, BAAI/bge-small-en, and fine-tuned models using hit rate metric and InformationRetrievalEvaluator